

The Hierarchy of Hate: Mixed Signals in the Combat against Hate Speech

VB verfassungsblog.de/the-hierarchy-of-hate-mixed-signals-in-the-combat-against-hate-speech/

Natalie Alkiviadou Di 6 Feb 2018

Di 6 Feb 2018

In May 2016, Facebook, Twitter, YouTube and Microsoft signed the 'Code of conduct on countering illegal hate speech online' with the European Commission. Since then, there has been a series of three monitoring cycles, during which public authorities and non-governmental organisations as cooperation partners of the Commission checked, over a period of several weeks, whether the companies are doing what they agreed to do. One of their central obligations is that they review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary. What is deemed illegal is defined by the Framework on Racism and Xenophobia, a document discussed further down. The last monitoring cycle took place between November 6th and December 15th 2017 in twenty-seven Member States (except Luxembourg). The results were comparatively favourable (overall) in comparison with the previous monitoring cycles. On average, IT companies removed 70% of the prohibited content, compared with a removal rate of 59% in the second monitoring cycle and 28% in the first.

On the face of it, this is good news. The European Commission is creating synergies with a variety of entities to crack down on hate speech, the removal rate of 'illegal hate speech' is rising and, thus, our pursuit (as a democratic society) of 'cleaning up' threats to pluralism, acceptance and solidarity is increasingly becoming efficient. However, it is not as simple as that. The Code of Conduct, as an initiative of the European Commission, impacts hate speech on social platforms in the European Union only. So, what about hate speech elsewhere, or outside social media, or beyond the four IT companies mentioned above, or even offline? Moreover, if international and European law had worked effectively enough to guide States (not private IT companies) in the right direction when it comes to hate speech regulation, would the liaison with private actors really be a necessity? Would watching, reporting, removing and then checking removals really be considered efficient in challenging online hate speech?

There is a number of varying thresholds to free speech regulation set out by relevant legal tools, namely Article 4 of the International Convention on the Elimination of all Forms of Racial Discrimination, Article 20(2) of the International Covenant on Civil and Political Rights, the EU's Framework Decision on Racism and Xenophobia and the Additional Protocol to the Cybercrime Convention. These variations, discussed below, can do nothing but confuse countries. Moreover, threshold or no threshold, anti-hate speech legislation developed on an international and European level has a major flaw, which goes against principles such as solidarity and equality. More particularly, apart from the justifiable direction of the ICERD, these frameworks are marred by what I refer to as the hierarchy of hate, namely the arbitrary focus on particular types of hate speech, such as racist speech, and the simultaneous disregard for other genres such as homophobic speech.

International Framework

Article 4, ICERD punishes, amongst others, all dissemination of ideas based on racial superiority or hatred and incitement to racial discrimination. Article 20(2) of the ICCPR prohibits (not punishes) any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. So, just on a United Nations level, we have both a threshold and a penalty discrepancy. Whilst the ICERD punishes the particular expression by law, the ICCPR simply prohibits it, therefore not necessarily calling on criminal law as the necessary tool. To confuse the situation more, the prohibited speech under the ICERD encapsulates a low threshold, including, for example, the dissemination of ideas based on racial superiority (with no necessity that these ideas constitute or call for hatred or violence). On the other hand, the ICCPR prohibits the actual advocacy for discrimination, hatred or violence. It is not, therefore, sufficient, for such speech merely to disseminate ideas of racial or religious superiority. In fact, the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression held that ‘the threshold of the types of expression that would fall under the provisions of Article 20 (2) should be high and solid.’

On a European level, we have two relevant documents, namely the Framework Decision on Racism and Xenophobia and the Additional Protocol to the Cybercrime Convention concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems. The first document constitutes the legal Framework of the Code of Conduct. Article 1.1(a), therein, holds that Member States shall take the measures necessary to ensure that the following intentional conduct is punishable:

‘publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin.’

On a Council of Europe level, the Additional Protocol to the Cybercrime Convention tackles racist and xenophobic material online. The lowest threshold prohibited conduct is the intentional and „without right“ dissemination of racist and xenophobic material through computer systems.

So, about thirty years down the line (after the two above UN documents), the most central European anti-hate speech document (the Framework Decision) heightens the threshold (we now need to demonstrate that conduct is intentional, that the expression is public and that it incites violence or hatred (not discrimination as in the case of Article 20(2) ICCPR). The threshold is further raised by Article 1.2 which holds that Member States may choose to punish only conduct which may disturb public order or which is threatening, abusive or insulting. So, if I am a country member of both the UN and the EU, I am definitely receiving mixed signals from the institutions I am part of as to what kind of expression should in fact be deemed illegal.

Hierarchy of Hate: An anathema to solidarity, equality and social justice

The ICERD deals with nothing but racist speech because it is a Convention about racism.

The ICCPR deals only with national, racial or religiously motivated hate speech but looks at no other types of prohibited expression such as homophobic or transphobic speech. Why? I cannot say. Maybe the issue is chronological, protecting sexual minorities was not as high on the agenda of the international community in the 60s, when the Covenant was drafted, as it is today. But, what I really cannot find a justification for is the EU's framework decision of 2008. Why was the focus only on hate directed against racial, religious, ethnic and related characteristics? What is the reason for the absolute disregard to other characteristics targeted by individuals, groups and institutions (particularly those related to a person's sexual identity and orientation)? What message is the EU sending with the choice to ignore the need for protection of these and other structurally and institutionally marginalised groups?

Intentional or not, the EU has demonstrated that it is only interested in criminalising particular types of hate speech whilst completely disregarding others. The same can be said for the Council of Europe, which chose an Additional Protocol to the Cybercrime Convention to focus only on racist and xenophobic material online. Note that the necessity of an Additional Protocol arose after the USA disagreed with the incorporation of its content in the Convention itself due to First Amendment considerations.

To conclude: The European Commission is trying to find ways to tackle hate speech on social platforms and has come up with the innovative strategy of (i) the Code of Conduct itself and (ii) the monitoring exercises. However, I fear that, notwithstanding the significance of these, they are just band aid approaches, and I wouldn't be surprised if the Commission knew this but chose to do something rather than nothing at all.

Let us not forget that the removal of hate speech on social networks relies primarily on the user whilst the frequency of material generated is hard to keep up with. What we need are concise socio-legal measures on both national and international levels to pull the roots of hate speech, many of which are nurtured by bad policies and wrong handling of perceived 'crises.' However, I am realist enough to see that this is probably a long shot. The least we could do, though, is to establish a well-rounded approach to regulating hate speech, with synergy amongst institutions, such as the UN and the EU, and one which eliminates the hierarchy of hate established to date.

LICENSED UNDER CC BY NC ND

SUGGESTED CITATION Alkiviadou, Natalie: *The Hierarchy of Hate: Mixed Signals in the Combat against Hate Speech*, *VerfBlog*, 2018/2/06, <https://verfassungsblog.de/the-hierarchy-of-hate-mixed-signals-in-the-combat-against-hate-speech/>.